



## Making and Using DNA Microarrays: A Short Course at Cold Spring Harbor Laboratory

David J. Stewart

*Genome Res.* 2000 10: 1-3

Access the most recent version at doi:[10.1101/gr.10.1.1](https://doi.org/10.1101/gr.10.1.1)

---

### References

This article cites 5 articles, 4 of which can be accessed free at:  
<http://genome.cshlp.org/content/10/1/1.full.html#ref-list-1>

### Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

---

To subscribe to *Genome Research* go to:  
<http://genome.cshlp.org/subscriptions>

---

# Making and Using DNA Microarrays: A Short Course at Cold Spring Harbor Laboratory

David J. Stewart<sup>1</sup>

*Meetings and Courses, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 USA*

**T**he conundrum is familiar. You are sent back in time to the Middle Ages with no artifact from the present, brought before the local ruler, and given 24 hours to prove you are indeed from the future, to impress the ruler and his advisors in some way, before you are executed in some suitably hideous fashion. What do you do?

Toying with this conundrum reveals how little we know in a practical sense about the everyday items that surround us. Can you fix your car *and* your computer? My guess is that few, if any, readers can do so. And so it was with some trepidation that Cold Spring Harbor Laboratory agreed to host a short course in the Fall of 1999, funded in part by the National Cancer Institute, in which students, primarily biologists, would not only print, use, and analyze DNA microarrays but would start the course by building the machines used to print the arrays. For some time, Patrick Brown and colleagues (Chu et al. 1998; DeRisi et al. 1997; Lashkari et al. 1997) at Stanford had been advocating the idea that smaller laboratories could enter the fray and hype surrounding these emerging microarray technologies by building machines rather than by buying them, a self-help philosophy that was strengthened by the Brown laboratory's web-based publication in June 1998 of the *MGuide*, a step-by-step guide to construct the arrayer, complete with parts list. Indeed, a number of laboratories have gone ahead and built their own machines.

Commercial vendors already offer some solutions for investigators interested in studying changes in genome-wide gene expression. Efforts by Steve

Fodor and others at Affymetrix (Santa Clara, CA) in the early 1990s had led to the development of the GeneChip technology, in which relatively costly photolithographic techniques are used to fabricate high-density microarrays of short sequences of single-stranded DNA base by base, but academic laboratories in particular find the technology both expensive and restrictive, the latter reflecting the fact that all of the arrays have to be manufactured by Affymetrix, presumably with a strong commercial perspective as to what genes (and from which species) are being arrayed. Because these arrays are composed of short (20-24 mer) oligonucleotides, they have application not only in monitoring "global" gene expression but also in the resequencing of genomic DNA, identification of single nucleotide polymorphisms (SNPs), and genotyping and will therefore have wide application in pharmacogenomics. But with current arrays sporting 40 features per gene (20 positive oligos designed to cover the length of the gene, and 20 mismatch controls containing identical sequences but containing a single centrally located mismatch), the Affymetrix approach can be considered to be overkill for many applications.

The second strategy revolves around commercialization of various aspects of the Stanford technique. For example, Synteni, a company that licensed the Stanford technology and was acquired by Incyte (Palo Alto, CA) in 1998, together with several other competitors, print their own arrays, where larger DNA fragments several hundreds of nucleotides in length are prepared by PCR in advance and then coated onto various flat substrates, primarily glass or nylon. These companies either sell arrays or array services, an approach that suffers

from similar restrictions to the Affymetrix approach in terms of which genes the companies decide to array. Many of these products consist of low thousands, hundreds, or even tens of arrayed sequences. Meanwhile, a third approach, midway between the second strategy and the purist Stanford approach, is to buy an arrayer from a commercial vendor such as Cartesian Technologies (Irvine, CA), and then make the DNA chips *de novo*. This offers flexibility to the investigator in terms of which sequences are arrayed, and the technical support of the vendor in case the printing robot breaks down or becomes unaligned—printing tens of thousands of discrete DNA "features" requires that these arrayers are tightly aligned in both horizontal directions. However, these arrayers have specifications no better and are currently at least twice the cost of home-built machines. This brings us back to the Stanford approach—build the machines from scratch. And to our own trepidation, could a group of 16 biologists—selected from a pool of >125 applicants on the basis of their biological interests rather than their machining skills—actually build the machines, albeit with expert guidance from members and former members of the Brown and Botstein laboratories in Stanford, such that they could be used to print high-density DNA microarrays (Table 1)?

As is usual for Cold Spring Harbor courses, the students included laboratory heads, senior scientists, and post-docs, plus two from Britain, and one each from Sweden, Germany, and New Zealand, with the remainder coming from academic laboratories in the United States with widespread interest in topics ranging from the cell cycle, origins of replication, cancer (and the development of anti-cancer vaccines),

<sup>1</sup>Corresponding author.  
E-MAIL [stewart@cshl.org](mailto:stewart@cshl.org); FAX (516) 367-8845.

**Table 1. Students and Faculty in the 1999 Cold Spring Harbor Laboratory Microarrays Course**

|                         |  |
|-------------------------|--|
| Jürg Bähler             | Cell Cycle Laboratory, Imperial Cancer Research Fund, London, UK                         |
| Arul Chinnaiyan         | Department of Pathology, University of Michigan Medical School, Ann Arbor, MI            |
| David Collingwood       | Department of Mathematics, University of Washington, Seattle, WA                         |
| Bruce Futcher           | Department of Microbiology, State University of New York at Stony Brook, Stony Brook, NY |
| Janet Hager             | Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT      |
| Christian Kaltschmidt   | Molecular Neurobiology Laboratory, University of Freiburg, Freiburg, Germany             |
| Thomas Kocarek          | Institute of Chemical Toxicology, Wayne State University, Detroit, MI                    |
| Maria Lagerstrom-Fermer | Department of Medical Sciences, University of Uppsala, Uppsala, Sweden                   |
| Matthias Lorenz         | Advanced Technology Center, National Cancer Institute, Gaithersburg, MD                  |
| Donald Love             | School of Biological Sciences, University of Auckland, Auckland, New Zealand             |
| Michele Marron          | Pathology, Immunology, and Experimental Medicine, University of Florida, Gainesville, FL |
| Vivek Mittal            | Cold Spring Harbor Laboratory, Cold Spring Harbor, NY                                    |
| Daniel Notterman        | Department of Molecular Biology, Princeton University, NJ                                |
| Michael Ryan            | Wadsworth Center, New York State Department of Health, Albany, NY                        |
| Arthur Thompson         | Department of Food Safety, Institute of Food Research, Colney, Norwich, UK               |
| Sudha Veeraraghavan     | Department of Molecular Biology, Scripps Research Institute, La Jolla, CA                |

Instructors: Ash Alizadeh (Stanford), Patrick Brown (Stanford), Max Diehn (Stanford), Michael Eisen (Lawrence Berkeley National Laboratory), Jo DeRisi (UCSF), and Paul Spellman (Stanford).

signal transduction, apoptosis and neurobiology. Preference was given to individuals whose applications strongly suggested that they would move swiftly to develop and apply this technology at their home institutions and make it available to other investigators. The explicit intention was to spread the application of these techniques as widely as possible, both geographically and scientifically.

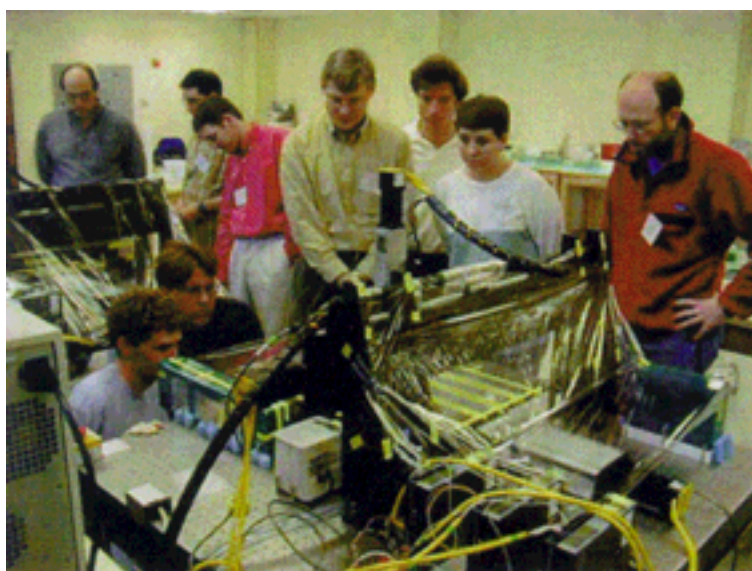
The students assembled at Cold Spring Harbor Laboratory on the night of October 19 to begin the 2-week course, and began building the arrayers the next morning. With one arrayer built in advance by Vishy Iyer and Jo DeRisi, a lead instructor in the course, serving as a guide, the students were able to build three complete machines by the third day of the course—these were long 16 hour days—despite “teething problems” in terms of broken or malfunctioning components (Fig. 1). Predictably, the students learned more from the problems that they encountered than an error-free assembly of the equipment might have offered.

By the fourth and fifth days, the course was printing duplicate arrays of the entire 6200-gene set of *Saccharomyces cerevisiae*, chips valued in excess of several tens of thousands of dollars by current commercial prices, using clones

provided by Stanford. With four machines in operation, the course laboratory, for 1 week in October at least, probably represented the largest chip printing facility anywhere, with a hypothetical annual capacity for >100,000–150,000 twenty-eight thousand spot arrays. The quality of these homemade microarrays may vary rather more widely than commercially available arrays, but with the cost differential so large between the two approaches, various kinds of error can be significantly

reduced by increasing the number of replicate arrays or even by altering the pattern of printing.

With sufficient arrays printed and available for experimentation, the students were ready to prepare samples for hybridization. Regardless of how DNA microarrays are fabricated, at this point methods for using these arrays start to coalesce, particularly in terms of gene expression analysis. Because of the enormous variation in the number of mRNA molecules being analyzed, and because

**Figure 1** Jo DeRisi (bottom left) and students examine the fully constructed arrayers.

of the complexities of the hybridization kinetics of individual DNA sequences, microarrays are used to measure the ratio between a reference and a sample, typically labeled with green and red fluorescent dyes, rather than the absolute quantity of transcript. It is for this reason that raw array data are typically represented as a grid of dots of varying intensities of red, yellow and green. The individual spot represents a marker for a given gene or sequence whereas the intensity of the red or green spots indicates the degree of expression difference between sample and reference but gives no information as to whether this is an abundantly or poorly expressed gene; bright yellow spots simply indicate good hybridization of equal numbers of red- and green-labeled molecules and imply no change in gene expression.

The students were able to use equipment loaned by various vendors to scan the slides and began the process of analyzing the data. One of the instructors, Michael Eisen, has been at the forefront of the development of a suite of freely available software tools, including ScanAlyze and Cluster, which help investigators work with the raw data (Fig. 2). Low-quality spots need to be identified, whether arising from inconsisten-

cies in the surface, poor printing, or poor hybridization. Spot intensity may vary across individual spots, and so various kinds of averaging have to be done, taking into account background signal. The clustering algorithms as developed by Eisen and others (1998) essentially allow the extraction of patterns of gene expression from a large quantity of data sets and use various strategies to help the investigator in visualizing large numbers of gene expression ratios. An elegant analogy used by these investigators to underscore the process is to take a Raphael painting, slice and dice the painting into thousands of randomly rearranged strips, and then attempt to reconstruct the original—one knows the pattern is there, but how does one (re)discover it? And the principal theme emerging from microarray experiments is that groups of genes that are functionally related tend to be coregulated at the transcriptional level.

Microarray technology has been criticized for diverging from the current trend for hypothesis-driven research. It strikes me that this is unfair—investigators using array technology seem to me to be the equivalent of the nineteenth century zoologists and botanists who traveled the world collecting everything they could

lay their hands on. Provided the collection is done well, the data are then available for others to study and draw inferences from. And if it accelerates the process of identifying genes of unknown function by virtue of their expression profiles, this is surely only a good thing. It is clear that as the amount of gene expression data and whole genomic information grows, it is vital that sufficient effort is spent on trying to develop ways in which data generated can be compared between laboratories. The challenges ahead lie as much in the development of sophisticated databases and advanced bioinformatics to mine reliable information from disparate data sets as in the relatively straightforward preparation of the arrays themselves.

The course allowed Brown, DeRisi, Eisen, and their colleagues to communicate their passionate conviction to a captive audience that arrays allow researchers both detailed and simultaneously holistic views of how organisms function. There is no doubt that we are going to witness a plethora of whole genome studies as the technology develops, and as more investigators begin to array not only DNA but antibodies and other proteins and to study noncoding DNA, transcription factor binding, protein-protein interactions and other macromolecular interactions. Although I cannot guarantee that any of these students would survive the judgement of the ruler in the familiar conundrum mentioned above, I am convinced that in the “new world” of microarrays (Brown and Botstein 1999), they will at least be capable of troubleshooting their arrayer (and, perhaps, their car and computer too).

## REFERENCES

- Brown, P.O., and D. Botstein. 1999. *Nat. Genet.* **21**: 33–37.
- Chu, S., J. Derisi, M. Eisen, J. Mulholland, D. Botstein, P.O. Brown, and I. Herskowitz. 1998. *Science* **282**: 699–705.
- DeRisi, J.L., V. Iyer, and P.O. Brown. 1997. *Science* **278**: 680–686.
- Eisen, M., P. Spellman, P.O. Brown, and D. Botstein. 1998. *Proc. Natl. Acad. Sci.* **95**: 14863–14868.
- Lashkari, D.A., J.L. DeRisi, J.H. McCusker, A.F. Namath, C. Gentile, S. Hwang, P.O. Brown, and R.W. Davis. 1997. *Proc. Natl. Acad. Sci.* **94**: 13057–13062.



**Figure 2** Michael Eisen explaining how to use his suite of software tools to discern underlying patterns of gene expression.